UBIQUITOUS RELAXATION IN BTI STRESSING—NEW EVALUATION AND INSIGHTS

B. Kaczer, T. Grasser\textsuperscript{1}, Ph. J. Roussel, J. Martin-Martinez\textsuperscript{2}, R. O’Connor, B. J. O’Sullivan, G. Groeseneken\textsuperscript{3}
IMEC, Kapeldreef 75, B-3001 Leuven, Belgium
\textsuperscript{1} CDL for TCAD, IµE, TU Wien, Austria; \textsuperscript{2} visiting from UA Barcelona, Spain; \textsuperscript{3} also ESAT Department, KU Leuven, Belgium
phone: +32 (16) 281-557, fax: +32 (16) 281-706, e-mail: Ben.Kaczer@imec.be

ABSTRACT

The ubiquity of threshold voltage relaxation is demonstrated in samples with both conventional and high-k dielectrics following various stress conditions. A technique based on recording short traces of relaxation during each measurement phase of a standard measure-stress-measure sequence allows monitoring and correcting for the otherwise-unknown relaxation component. The properties of relaxation are discussed in detail for pFET with SiON dielectric subjected to NBTI stress. Based on similarities with dielectric relaxation, a physical picture and an equivalent circuit are proposed.

INTRODUCTION

Bias-temperature instabilities (BTI) in CMOS technologies with advanced gate stacks are currently at the forefront of reliability concerns. Accurate evaluation of BTI degradation, necessary for correct lifetime extrapolation, is proving far from straightforward. On one hand, on-the-fly techniques are hampered by experimental difficulties with recording the initial (time = 0 s) state of the stressed device under test (DUT) and with deconvoluting FET parameter (mobility and threshold voltage $V_{th}$) shifts [1]. On the other hand, standard measure-stress-measure (MSM) techniques used by most in industry, give hard-to-compare results due to ignored or unknown relaxation of the measured parameters, such as the FET threshold voltage shift $\Delta V_{th}$.

The $\Delta V_{th}$ relaxation is typically observed to span many decades in time, from microseconds or possibly less [2,3] to days (Fig. 1) [4-6] indicating an underlying dispersive mechanism [7]. In an early attempt to circumvent the problem of $V_{th}$ relaxation, we have designed a technique to speed up the evaluation of $\Delta V_{th}$ [6,8]. This measurement technique, now considered for a JEDEC standard, is based on a rapid, single-point measurement of $I_{D,lin}$ at $\sim V_{th}$. As is apparent from Fig. 1, however, achieving full correction of $V_{th}$ relaxation with this technique is literally an uphill battle. This is because typical standard semiconductor analyzer equipment, capable of taking a current measurement in ~0.1 s, will only measure an unknown fraction of the recovery already in full progress. Extrapolation of thus-obtained data is then less than meaningful. Circumventing this problem is potentially possible with ultra-fast measurement setups [9], which at the moment, however, are still highly custom-specific and used by a few research groups only.

In this work we focus on MSM measurements with commonly-available off-the-shelf semiconductor analysis equipment. We show that far more insight into the BTI processes can be gained by abandoning the as-fast-as-possible approach and instead recording a short portion of recovery during every measurement phase. Then, by making an assumption about the behavior of BTI recovery [10,11], these short individual traces can be fitted together and the recoverable $R$ and permanent $P$ components of BTI degradation can be separated, independently evaluated and extrapolated to operating conditions.

The subject of lifetime extrapolation per se is not elaborated on in this work [12,13]. Here we underline the importance of $\Delta V_{th}$ relaxation (i.e., the $R$ component) on lifetime projection and we discuss its properties. Using the described measurement technique we have found that the relaxation of $\Delta V_{th}$ is prevalent in all studied systems so far (e.g., SiON pFET and nFET, NBTI and PBTI, AC NBTI, high-k pFET NBTI, substrate hot-carrier stress, etc.). Furthermore, based on more extensive evaluation of relaxation in ultrathin SiON pFETs, we show that $\Delta V_{th}$ relaxation has some properties similar to dielectric relaxation [14,15]. We emphasize that these similarities are purely phenomenological, stemming solely from both mechanisms being dispersive in nature. However, we show that lessons learnt from the analogies could be useful both for circuit simulation of $\Delta V_{th}$ relaxation and its microscopic modeling.

MEASUREMENT TECHNIQUE

Experimental considerations and technique description

Standard FETs with preferably short gate length $L$ (~0.13 to 0.5 μm) and wide gate width $W$ (~10 μm) to maximize FET current are used. The DUT is biased with small $V_D$ (typically ±50 mV) and $V_S$ = 0. If the measurement setups [9], which at the moment, however, are still highly custom-specific and used by a few research groups only.

![Fig. 1: Relaxation of $\Delta V_{th}$ in pFET with 1.4 nm SiON, measured as illustrated in Figs. 2b, c, and d. The logarithmic transient spans 8 decades and most likely starts already well below 1 ms. Conventional equipment capable of taking a reading in ~0.1 s measures the sum of an unknown fraction $P$ of the recoverable component $R$ on top of the permanent degradation component $P$. Any extrapolation using thus extracted $\Delta V_{th}$ is questionable.](image-url)
0V (Fig. 2c). This ensures that the stress voltage applied across the gate dielectric is approximately uniform. As in the case of the “fast” NBTI evaluation technique (Fig. 2a) [6], an initial I-V\textsubscript{G} characteristic (Fig. 2d) is first recorded to obtain a conversion “table” for all subsequent FET current (I\textsubscript{S} or I\textsubscript{D}) measurements. To minimize any unaccounted stressing of the DUT, care should be taken to measure this I-V\textsubscript{G} curve only up to ~V\textsubscript{meas}. The value of V\textsubscript{meas} is typically chosen around ±50—100\times W/L nA so that the measurement equipment can discern fractional changes in this current [16].

The DUT is then subjected to a preprogrammed MSM sequence of gate voltages V\textsubscript{G}, comprising of alternating stress phases at V\textsubscript{stress} and measure (or “relaxation”) phases at V\textsubscript{meas} (Figs. 2a and b). Small V\textsubscript{G} remains applied and the FET current is recorded during the entire sequence. The additional data collected during the stress phases are equivalent to on-the-fly measurements for the respective phases and can provide further information about the DUT behavior [17]. We also found it beneficial to monitor the initial FET current for ~100 s before V\textsubscript{stress} is applied (see Fig. 2b) to ascertain the DUT is stable at V\textsubscript{meas}. As is typical in most BTI experiments, the duration of stress phases is geometrically increased to cover several decades.

The new V\textsubscript{G} waveform (Fig. 2b) differs in the design of the measure phase. In the previous “fast” BTI evaluation technique, each measure phase was performed as fast as possible (~0.2 s on HP4156) and typically only one current point was recorded (see Fig. 2a). In the new measurement scheme (Fig. 2b), each measure phase is designed to collect maximum amount of information about relaxation in the short interval. A fixed time of ~12 s is used in this work; a fixed fraction (e.g. 1/100) of the next stress phase duration is also possible [13, 16].

Direct-to-buffer functionality of grouped Keithley 2602 digital multi-meters (DMM’s) is employed in our setup to capture data at short times. The data-sampling rate is sequentially reduced (“r1”, “r2”, “r3” in Fig. 2b) by a script running directly on the DMM to cover the 12 s interval approx. uniformly on a logarithmic scale, resulting in efficient use of the DMM memory. Long-integration time DMM measurements with exponentially distributed delays (“log” in Fig. 2b) are added to the fast-buffer sequence for long relaxation measurements, typically the last measure phase of the MSM sequence.

The measurement sequence is concluded with a final I-V\textsubscript{G} characteristic to check the DUT. Finally, the FET current data recorded during the MSM sequence are converted to ΔV\textsubscript{th} using the initial I-V\textsubscript{G}, as illustrated in Fig. 2d [6].

The entire experiment (Keithley 2602 DMMs, Süss PA300 probe station, thermo chuck) is in our case controlled over GPIB from a PC using a framework of perl script subroutines. The use of high level programming language facilitates simple tasks, such as reformating low-level GPIB data and processing results, but also allows e.g. preprogramming long experiments on many DUTs and even making decisions (e.g. “skipping” faulty devices) while running unattended.

Measurement technique advantages

Once the FET has been exposed to BTI stress, its out-of-equilibrium parameters will relax over many time scales, from microseconds (or less) to days (or more). These distributed-time-scales parameter transients cannot be therefore “waited out” during reasonable time (see Fig. 1). Due to the same property of the relaxation component, also the “obvious” solution to accelerate relaxation before sensing is far from trivial. As it is shown in Fig. 3, relaxation can be accelerated by applying opposite (positive for pFET) voltage after stress is removed. Even after such acceleration, the relaxation takes 10\textsuperscript{3} – 10\textsuperscript{4} s to approach within 10% of the final P value [13]. Choosing the acceleration voltage is also problematic—applying too high opposite voltage results in BTI stress and defect creation of its own (see the V\textsubscript{relax} = +2 V curve in Fig. 3).

![Fig. 2: (A) Previous “fast” and (B) the new BTI evaluation techniques. Unlike the “fast” technique (A), each “measure” phase is now designed to collect maximum information about relaxation in a short interval (B). “r1”, “r2”, “r3”, and “log” represent different sampling rates to efficiently cover logarithmic time scale. (C) DUT bias during measurement and (D) an illustration of the post-stress FET current-to-ΔV\textsubscript{th} conversion.](image-url)
Before we discuss experimental results obtained using the described technique, namely the properties of the relaxing component, we briefly summarize possible experimental caveats and drawbacks of the technique.

Firstly, it is essential that the technique is only applied to devices showing stable FET current when \( V_{\text{meas}} \) is applied before the application of stress. This is checked by monitoring each DUT at \( V_{\text{meas}} \) for 100 s (see Fig. 2b).

The pivotal assumption of the measurement technique is that at \( V_{\text{meas}} \), all changes in the recorded FET current are due to \( \Delta V_{th} \) only. Choosing \( V_{\text{meas}} \) = \( V_S \) minimizes influences of both subthreshold slope and mobility changes [1]. Other factors potentially influencing the FET current during measurement include changes in gate oxide leakage (SILC and soft breakdown) and in drain junction leakage (since \( V_D \neq 0 \)). The latter issue, occurring e.g. in FETs with small-band gap Ge substrates [19] can be reduced by monitoring the source current, since \( V_S = V_D = 0 \). Presence of all other current perturbations should be checked by comparing the shapes and magnitudes of all currents of the initial and final I-\( V_G \) characteristics.

Finally, it is crucial to check the actual waveform (Fig. 2b) applied on the gate of the DUT with an oscilloscope. For example, some measurement setups can momentarily dwell at 0V during each voltage transition. The degradation at measure voltage then appears smaller due to the inadvertently accelerated relaxation, as illustrated in Fig. 3.

**RESULTS AND DISCUSSION**

**Examples of relaxation ubiquity**

Fig. 5 shows examples of data collected on both SiON and high-k samples under various stress conditions. In all cases, relaxation of \( \Delta V_{th} \) is apparent, underlying the need to record and evaluate the recovery in detail. Also note that on all samples the (final topmost) relaxation is observed to stretch over 6 decades in time, with no apparent sign of saturation at either the short or long relaxation times. In contrast to that, the 1-exp(-t/\( \tau \)) discharge of traps with a single \( r \) occurs over ~1.4 decades in time (assuming transition from 10% to 90%), while the Reaction-Diffusion (RD) model [20] relaxation is expected to occur over ~3.8 decades in time (10%—90%) [10]. From this we conclude that in all cases in Fig. 5 the relaxation is most likely controlled by a dispersive mechanism with widely distributed time scales [7, 11].

We have noticed that while the relaxation is always present, its relative magnitude compared to the overall degradation can vary among samples and stress conditions. While the relaxation is dominant after both NBTI and PBTI in pFETs with both SiON and high-k gate oxides, in some cases, such as hot channel electron injection (HCI) in SiON nFET’s, it could be less pronounced (not shown). In such a system [21], relaxation can likely be ignored in the first approximation.

**Properties of the recoverable component**

The ubiquitous recoverable component is studied in more detail on pFET’s with 1.4 nm EOT SiON dielectrics subjected to NBT I stress (unless otherwise noted). In particular we discuss its following properties: universality and relaxation time, stress time, voltage, temperature, and AC stress duty cycle dependences.

**Relaxation time dependence—universality of relaxation**

Previously we have noticed [6,13] that the long, log-like BTI relaxation transients can be well described by
discussed in Ref. [22]. The universality of BTI relaxation was also observed and the universal relaxation time parameter [7].

The assumption of universality of BTI relaxation allows us to fit all short relaxation transient “snapshots” measured during each measure phase with a single set of parameters $B$ and $\beta$ for each MSM sequence [13]. The fitting procedure further assumes

$$\Delta V_{th}(t_{\text{relax}}, t_{\text{stress}}) = R(t_{\text{stress}}, t_{\text{relax}}) + P(t_{\text{stress}}),$$

and

$$R(t_{\text{stress}}, t_{\text{relax}}) = R(t_{\text{stress}}, t_{\text{relax}} = 0) r(\xi).$$

In Eq. 3, $R(t_{\text{stress}}, t_{\text{relax}} = 0)$ represents the “full” recoverable component extrapolated to $t_{\text{relax}} = 0$.

The fits for a diverse combination of samples and stressing conditions are shown in Fig. 5. In all the cases, Eq. 3 fits the relaxation data well. $\beta$ is typically between 0.14 and 0.20 for all pFET samples (both SiON and Hf-based high-k dielectrics) subjected to NBTI stress.

While we observe the universality of relaxation in most production-quality gate stacks (with either SiON or Hf-based dielectrics on Si substrates), we have noticed that relaxation does not fully follow Eq. 3 in some very experimental gate stacks, such as those based on Ge substrates [19], dielectrics containing Dy and La, etc. [23,24]. In those systems, an additional mechanism, such a strong trap level in the dielectric, is suspected to be active during relaxation. Recording the relaxation transients in such cases is still beneficial, as it exposes processes in the sample that would be otherwise overlooked during standard BTI analysis [24].

**Stress time dependence**

The above-described fitting procedure is now applied to extract the recoverable component for every measure phase of the NBTI MSM sequence. The semi-log plot in Fig. 6a shows $R(t_{\text{stress}}, t_{\text{relax}} = 0)$ as function of the cumulative stress time $t_{\text{stress}}$. $\Delta V_{th}(0)$ appears to scale as $\log(t_{\text{stress}})$ [2,22].

We note that because of the relatively small increase of $\Delta V_{th}/V_{\text{meas}}$ in the measured range of $t_{\text{stress}}$, the same data will also appear to follow a power law in a log-log plot, however, with a very small exponent of 0.04—0.06. This is consistent with $\Delta V_{th}$ data extracted in some ultra-fast measurements [9], which observe exponents < ~0.1. Ultra-fast measurements are able to capture a much larger part of the relaxing component (see Figs. 1 and 4) and consequently, the measured $\Delta V_{th}$ dependence will be dominated by this component.

**Voltage dependence**

Voltage dependence of the recoverable component is studied by plotting $R(t_{\text{stress}} = 6000 s, 0)$ from Fig. 6a versus the $V_{\text{stress}} - V_{\text{meas}})$, i.e. the voltage increase applied during stressing. Since $V_{\text{meas}}$ was chosen ~ $V_{th}$ of the fresh device (see above), $V_{\text{stress}} - V_{\text{meas}}$ is approximately proportional to the electric field in the oxide during stress. Scaling of the BTI mechanism with the oxide electric field has been generally accepted [25]. In Fig. 6b it is observed that the dependence of $R(t_{\text{stress}}, 0)$ on this quantity is superlinear, close to the second power [26].
Temperature dependence

\( R(t_{\text{relax}} = 0) \) is weakly temperature dependent, with activation energy \( \sim 80 \text{ meV} \), as shown in Fig. 7. Interestingly, the “dispersion parameter” \( \beta \) (extracted at different temperatures, not shown) is seen to be independent of temperature, which contrasts with the predictions of dispersive hydrogen transport models [6,10].

---

AC stress duty cycle dependence

Although the properties of degradation after constant (static or DC) BTI stress are important, they are of limited use for digital CMOS applications with continuously switching FET devices. It has been generally observed that dynamic (i.e., AC) unipolar NBTI stress with 50% duty cycle results in about half degradation of the equivalent static stress [20]. The dependence on the duty cycle (called duty factor \( \text{DF} \) here), however, has been seldom studied. In Ref. [27] we have reported for the first time a \( \Delta V_{\text{th}} \)-DF dependence with an inflection point around DF \( \sim 50\% \). Qualitatively identical dependence, re-measured with the new MSM technique is shown in Fig. 8a. We argue that this distinctive shape in Fig. 8a is a fundamental feature of BTI relaxation and should serve as a touchstone for all proposed NBTI models.

Fig. 8b shows the full recoverable \( R \) and the permanent \( P \) components extracted from data in Fig. 8a. As can be seen, the particular shape of the DF dependence is chiefly due to the recoverable component \( R \).

We also note a peculiar property of Eq. 1—when applied on the microsecond scale, it qualitatively explains the \( R \)-DF dependence in Fig. 8b [28]. Assuming (see inset of Fig. 8b)

\[
DF = \frac{t_s}{t_s + t_r} = \frac{1}{1 + \frac{t_r}{t_s}},
\]

where \( t_s \) and \( t_r \) are now the stress and relaxation times within each cycle, we obtain

\[
DF = \frac{1}{1 + \xi}
\]

and hence,

\[
\xi = \frac{1}{DF} - 1.
\]

Fig. 8b shows the excellent fit of \( \text{const} \times r(1/\text{DF}-1) \) to the extracted \( R(t_{\text{relax}} = 6000 \text{ s}, 0) \) component.

Curie-von Schweidler-like behavior of BTI relaxation

The ubiquitous recoverable component is studied in more detail by recording \( \Delta V_{\text{th}} \) recovery for 8 decades of time. As can be seen in Figs. 9a and b, both the continuous and the pulse-like stress result in a logarithmically decreasing \( \Delta V_{\text{th}} \) over the full time range.
Implications for physical modeling

Charge trapping and interface state creation (due to hydrogen depassivation) are the mechanisms most frequently invoked to explain BTI. Relaxation of the BTI degradation, i.e., the recoverable component $R$, is then typically ascribed to interface state hydrogen repassivation [20] or to neutralization of trapped charge [30], while the permanent damage has been related to unpassivated interface states [30] and to deep hole traps [31,32]. This discussion implies it is theoretically possible that i) two independent mechanisms are respectively responsible for the $R$ and $P$ components, but also that ii) a single mechanism is responsible for both degradation components. The microscopic picture suggested in Fig. 10b provides a new impulse for physical modeling of BTI phenomena in general, but also for investigating the possibility of a single common cause for both degradation components [17].

![Fig. 10: The behavior of the 1st derivatives of relaxation transients in Fig. 5, also known as the Curie-von Schweidler law, suggests applying similar (a) circuit and (b) microscopic models to BTI relaxation.](image)

Implications for electrical modeling

In order to study the reliability of CMOS circuits, an equivalent circuit for a FET emulating the effect of an arbitrary BTI sequence is needed. In its simplest form, the permanent and recoverable parts of the threshold voltage shift can be introduced as shown in Fig. 11a. We find that most properties of the recoverable component discussed above can be reproduced when ohmic resistors in each RC element in Fig. 10a are replaced with a non-linear component (simulated by 2 diodes with different parameters, see Fig. 11b). While this non-linear component is fixed for all elements, the capacitor values are varied by a factor of 10 from element to element. 20 elements are used in the calculation, carried out in the C language for speed, but cross-checked in the ADS circuit simulator. The threshold voltage shift at any moment is assumed to be proportional to the simple average of voltages on all capacitors, i.e., no weighting of elements is assumed for simplicity.

The equivalent circuit qualitatively correctly reproduces the logarithmic relaxation after BTI stress (Fig. 12; cf. Fig. 1), logarithmic increase during stress (Fig. 13a; cf. Fig. 6a), and even the superlinear voltage dependence (Fig. 13b; cf. Fig. 6b).

![Fig. 12: Calculation with the equivalent circuit in Fig. 8b correctly reproduces (a) the log-like transients of the recoverable component correctly reproduced for both DC ($DF = 100\%$) and AC ($DF < 100\%$) BTI stress (cf. Fig. 1).](image)

![Fig. 13: Calculation of the recoverable component as per Fig. 8b correctly reproduces (a) the log-like increase with stress time (cf. Fig. 6a) and (b) the superlinear dependence on stress voltage (cf. Fig. 6b).](image)
conditions. A MSM technique based on recording short traces of conventional and high-k dielectrics and subjected to various stress [2] H. Reisinger, O. Blank, W. Heinrigs, A. Mühlhoff, W. Gustin, [1] T. Grasser, P.-J. Wagner, Ph. Hehenberger, W. Gös, and B. Fig. 8) for AC BTI stress. Note also that the value at reproduces the distinctive dependence on the duty factor (Fig. 14; cf. about half of the degradation at DC (\(DF = 100\%\)) for all relaxation times. This could explain why approximately \(\frac{1}{2}\) the degradation in AC stress wrt DC stress is commonly observed by most researchers, independently of the measure (i.e., relaxation) time.

**CONCLUSIONS**

\(\Delta V_{th}\) relaxation is shown to be prevalent in FETs with both conventional and high-k dielectrics and subjected to various stress conditions. A MSM technique based on recording short traces of \(\Delta V_{th}\) relaxation is introduced, exposing processes in the DUT that would be otherwise overlooked during standard MSM BTI analysis. In samples showing universality of relaxation, this property is used to correct for the otherwise-unknown \(\Delta V_{th}\) relaxation component. Properties of \(\Delta V_{th}\) relaxation following NBTI stress of pFETs with SiON gate dielectric are discussed in detail. Based on similarities with dielectric relaxation, an equivalent circuit and a physical picture are proposed for \(\Delta V_{th}\) relaxation.

**ACKNOWLEDGEMENTS**

This work has been partially supported by the Spanish MCyT (TEC2004-00798/MIC).

**REFERENCES**


